

动物基因组学重测序的应用研究进展

汪文强^{1,2}, 赵生国², 马利青³, 郭继军⁴, 马月辉^{1*}, 赵倩君^{1*}

(1. 中国农业科学院北京畜牧兽医研究所, 北京 100193; 2. 甘肃农业大学动物科学技术学院, 兰州 730070;

3. 青海省畜牧兽医科学院, 西宁 810016; 4. 青海省畜牧总站, 西宁 810001)

摘要: 随着第二代测序技术的研发和应用, 基因组学的研究不断出新, 为其带来了更新的科研方法和解决方案。基因组测序可以更深入地了解一个物种的分子进化、基因组成和基因调控等特点, 特别基因组重测序技术的发展和运用, 将基因组学的研究推向了多领域、多样化、多功能的新阶段。现已从变异检测、性状定位、遗传图谱构建、群体进化分析等方面取得丰硕成果。文章阐述了动物基因组重测序学领域中全基因组测序技术和简化基因组测序技术的应用现状和发展趋势。

关键词: 重测序; 群体进化; 变异检测; 性状定位; 遗传图谱

中图分类号: S813.3

文献标志码: A

文章编号: 0366-6964(2016)10-1947-07

The Research Progress and Application of Resequencing Based on Animal Genomics

WANG Wen-qiang^{1,2}, ZHAO Sheng-guo², MA Li-qing³, GUO Ji-jun⁴, MA Yue-hui^{1*}, ZHAO Qian-jun^{1*}

(1. *Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China;*

2. *College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China;*

3. *Qinghai Academy of Animal Science and Veterinary Medicine, Xining 810016, China;*

4. *Animal Husbandry Station of Qinghai, Xining 810001, China)*

Abstract: With the application and development of the next generation sequencing technique, the researches of genomics are constantly updating, which finds out new solutions and technologies to genomics. The genome sequencing is competent to learn the population evolution, gene composition and gene regulation deeply, especially the application and development of genome resequencing technology, which makes the genome research come into being a new era in multiregion, diversification and multifunction. Nowadays the next generation sequencing technique has made a large progress in mutation detection, fine mapping of important genes, genetic map construction, analysis of population evolution, and so on. The review states application status and development tendency of whole genome sequencing technology and reduced-representation genome sequencing technology in animal genome resequencing.

Key words: resequencing; population evolution; mutation detection; fine mapping of important genes; genetic map

随着 Sanger 测序技术的限制性, 第二代测序技术(Next generation sequencing, NGS)的优势逐渐凸显, 对重测序技术的发展起到了重要的作用。

NGS 的核心思想是边合成边测序, 即通过捕捉新合成的末端的标记来确定 DNA 的序列, 测序技术成本低、高通量、快速、高效等特点能有效地鉴别单核

收稿日期: 2015-12-30

基金项目: 中国农业科学院科技创新工程(ASTIP-IAS01); 国家自然科学基金项目(31201765); 国家绒毛用羊产业技术体系(CARS-40-01)

作者简介: 汪文强(1991-), 男, 甘肃天水人, 硕士生, 主要从事动物遗传育种与繁殖研究, E-mail: 187931128479@163.com

* 通信作者: 赵倩君, 副研究员, E-mail: zhaoqianjun@caas.cn; 马月辉, 研究员, E-mail: yuehui.ma@263.net

核苷酸多态性 (SNPs) 标记、插入和缺失 (InDel) 标记^[1-3]。基因组重测序是指基于第二代测序,也可以是第一代的,对之前的测过序的基因组再测一遍,并对个体或者群体样品进行分析并获得需要的信息。基因组序列的完整性和有效性有利于全基因组重测序,以及之后通过比较基因组序列和全基因组重测序序列,可以在全基因组范围发掘 SNP 和 InDel 标记^[4-5],同时也有利于基于简化基因组重测序的 SNP 和 InDel 的研究。随着第二代测序技术的快速发展,重测序技术在动物变异检测、基因挖掘、性状定位、遗传进化等多领域均取得了丰硕的成果,得到了充分的发展与应用。

1 全基因组重测序

全基因组重测序是对已知基因组序列的物种进行不同个体的基因组测序,并在此基础上对个体或群体进行差异性分析,进而找到大量的单核苷酸多态性位点 (SNP)、插入和缺失位点 (InDel) 以及结构变异位点 (SV) 等变异信息。全基因组重测序技术可广泛用于变异检测、遗传图谱构建、性状定位和群体进化研究。

1.1 群体进化

Y. Qu 等对地山雀进行全基因组测序,并对其近缘种大山雀、黄颊山雀及黑尾地鸭进行重测序,发现地山雀与大山雀和黄颊山雀大约在 7.7 百万~9.9 百万年前产生了分化,从全基因组水平明确了地山雀的分类问题,同时在高原适应性遗传机制上分析发现免疫基因、尤其是和 MHC 相关的基因家族发生了显著收缩或者丢失;地山雀低氧适应及能量代谢相关的基因发生了快速进化;地山雀骨骼发育相关的基因发生了快速进化;嗅觉相关的基因家族也发生了显著收缩^[6]。M. Li 等对 6 个代表性藏猪群体、5 个四川盆地特有猪种共 48 个样本进行全基因组重测序(测序深度达 131X)并结合 55 个欧亚野猪及家猪的基因组数据进行群体遗传学分析,最终在藏猪中鉴定出低氧适应、能量代谢等共 268 个适应高原环境的快速进化基因,揭示了藏猪高原适应性的遗传机制;通过比较基因组学分析发现藏猪和家猪的祖先可能早在 690 万年前就开始向不同方向进化,分化时间可能早于牦牛和家牛(490 万年前)以及人类和黑猩猩(500 万~700 万年前);另外,对野生种和驯养种的分析表明,与自然选择相比,人工选择可更有效的塑造驯养动物基因组;欧亚猪种

存在明显的遗传背景差异,欧亚地理隔离造成的遗传结构差异甚至超过了野生和驯化的差异^[7]。Q. Xia 等利用重测序技术研究了家蚕的驯化事件和驯化基因,完成了 40 个个体的重测序,每个个体测序深度大约为三乘,覆盖基因组序列的 99.88%,共鉴定出 16 000 000 多个 SNPs;研究发现,家蚕的基因明显不同于野生型,但它们保持着较大级别的遗传变异,这表明短期的驯化事件包含了大量的个体;研究还发现,在驯化过程中有重要作用的 354 个候选基因存在选择信号,其中一些基因在丝腺、肠道和睾丸中过表达。这些数据对家蚕的驯化事件起到了重要的解释作用^[8]。X. Zhou 等利用全基因组测序技术研究了滇金丝猴的进化史,共确定了 21 813 种编码蛋白的基因,其中 89.7% 得到了转录组验证;根据这些编码蛋白的基因,研究人员绘制了灵长类动物进化的时间表;滇金丝猴和猕猴分别有大约 3 054 和 3 311 个特异性 LINE1 插入,其超过了人类(2 365 个新的插入)和黑猩猩(1 841 个新的插入),然而早期猴的系谱特异性 SVA 较低于类人猿。在早期猴跟类人猿的比较中,未发现基因家族扩展或收缩事件有显著性差异。总体讲,在早期猴的 ALU 富集并未导致全基因组范围类基因家族结构显著性的重排和改变^[9]。

1.2 变异检测

R. Zhang 等将人乳铁蛋白 (Human lactoferrin, hLF) 表达相关基因导入牛基因组中,应用二代测序技术检测 3 头转基因牛(2 头 F0 代,1 头 F2 代)外源插入 DNA 的分子特性、插入位点和整合拷贝数,发现外源插入 DNA 上不同区域在样本基因组上存在 2~8 倍的拷贝数变异,表明外源 DNA 在转基因物种基因组上发生了复杂的重排事件^[10]。D. W. Seo 等采用二代测序技术对韩国本地鸡种的 SNPs 进行分析,研究从 29 号染色体和 Z 染色体上共获得 $(4\,006\,068 \pm 97\,534)$ 个 SNPs,在这些已鉴别的 SNPs 中已知和新发现的 SNPs 分别为 $(2\,948\,648 \pm 81\,414)$ 个和 $(1\,047\,951 \pm 14\,956)$ 个,其中已知但未被定义的 SNPs 有 $(1\,181 \pm 150)$ 个,新出现未被定义的 SNPs 有 $(8\,238 \pm 1\,019)$ 个,同义 SNPs、错义 SNPs 和特征改变了的 SNPs 分别是 $(26\,266 \pm 1\,456)$ 、 $(11\,467 \pm 604)$ 和 $(8\,180 \pm 458)$ 个^[11]。P. D. Keightley 等对黑腹果蝇进行全基因组重测序以推测自发突变的发生率及特性。该研究利用 3 种突变集聚型样本,以这 3 种样本的基因组

DNA 为模板对其进行全基因组测序,研究人员绘制了包含 174 个单核苷酸突变位点的图谱,并计算了单核苷酸突变的发生几率,通过测序发现没有任何的假阳性,研究还发现了很多 G/C 到 A/T 的突变及 A/T 到 G/C 的突变,且 G/C 到 A/T 的突变是 A/T 到 G/C 突变的两倍,编码区与非编码区的 SNP 突变率没有显著性差异^[12]。C. J. Rubin 等利用全基因组重测序鉴别家鸡在进化过程中可能存在的突变。对家鸡及其野生祖先红原鸡进行了全基因组重测序,共发现了 7 000 000 个 SNPs, 1 300 个缺失突变和若干选择性清除,并且发现 *TSHR* 基因在该研究的所有家鸡中均存在选择性清除,该基因在脊椎动物的代谢调节及生殖过程中发挥着重要作用^[13]。K. T. Lee 等通过全基因组重测序技术研究了韩牛纯合性区域,确定了 4 700 000 个 SNPs 和 4 000 000 个 Indels。在 8 360 个基因中检测到了大约 25 000 个非同义 SNPs 位点、剪接位点变异体、以及编码插入缺失(NS/SS/IS)。此外,还发现与肉质和抗病性状相关基因 25 个。研究结果将为与牛重要经济性状相关基因的随意突变或基因鉴定提供科学的数据^[14]。B. Zhan 等通过基因组重测序和高通量基因型分型综合评估了牛基因组的变异。研究发现插入和缺失位于邻近蛋白的 N 末端和 C 末端,且 3 倍大小富集。研究结果提供了高分辨率不同类型的牛基因组变异,并证明结构变异作为基因组变异的主要成分超过序列变异^[15]。G. Yi 等基于全基因组重测序基础研究 12 只来源于不同品种鸡的全基因组 CNV。共发现了 8 840 个 CNVs 区域,覆盖了 98.2 Mb,代表了鸡基因组的 9.4%,这些 CNVs 大小不一,从 1.1~268.8 kb 不等,平均长度为 11.1 kb。总共预测到 2 214 个 CNVs 跨越 2 216 个具有特定生物学功能相关的 RefSeq 基因。同时被部分 CNVs 覆盖的区域发现了 *FZD6 L* 基因和 *IMS1* 基因,这两个基因与疾病易感性和抗病性相关^[16]。

1.3 性状定位

E. Axelsson 等应用第二代测序技术,对分布于世界各地的 12 匹狼和 60 条狗(14 个品种)进行了全基因组重测序(狼个体测序深度 6.2X,狗基因组混池测序深度 29.8X),探究狗对淀粉类食物适应性的关键性基因,研究共发现了 3 786 655 个 SNPs, 506 148 个 Indels 和 26 619 个 CNVs,分析表明,有 10 个基因与淀粉的消化和脂肪代谢明显相关,结果

表明,相对于肉食性的狼而言,狗对淀粉特异的适应性使得狗的祖先能很好的从富含淀粉类的食物中获得能量并繁衍生息,这对早期狗的驯化起到了重要作用^[17]。M. E. Bowen 等针对斑马鱼基因组大、缺乏自交系的特点,有效地使用新的并行测序技术绘制并检测了位于斑马鱼中的突变。全基因组测序利用了大量的参考 SNP 数据库通过覆盖率低来定义纯合逐步下降的区域,其全基因组测序的 DNA 池来源于有限个 F2 突变体,通过该方法,绘制了每一个不同的斑马鱼突变体图,在剩下的个体中研究人员选择了两个个体,并测序确定了可能诱发的错义突变和候选基因突变。此外,在一个已鉴定的突变体中证实 *bmp1a* 中存在一个错义突变^[18]。J. D. Merker 等采用全基因组重测序技术分析研究了早期原发性骨髓纤维化患者低突变和非复发性突变的候选基因,结果发现,原发性骨髓纤维化细胞基因组体细胞突变率很低,与造血器官肿瘤基因组中的结果一致。全基因组 DNA 测序与 RNA 表达数据相结合,鉴定出 3 个具有潜在功能的体细胞突变:*CARD6* 同义突变、*BRD2* 基因 5' 非翻译区 19 个碱基的缺失和 *KIAA0355* 非同义突变,另外还鉴定出 *CAP2*、*SOX30* 和 *MFRP* 3 个基因发生了基因突变,这 6 个基因在 178 个病人样本中得到了验证^[19]。H. D. Daetwyler 等利用全基因组重测序技术,研究了 2 头奶牛和 232 头公牛的群基因组,测序深度为 8.3 倍覆盖率,这些群体种含来自荷兰种 129 头,泽西种 15 头以及德国种 43 头。研究共鉴定了 2 830 000 个变异,每 1 000 个碱基上平均含有 1.44 个杂合位点。研究发现了与产奶、胚胎死亡、骨骼畸形以及卷毛相关的基因,并得出牛繁殖力下降的主要原因跟胚胎死亡有显著关系。该项研究为提高产奶和产肉提供了科学依据^[20]。

1.4 遗传图谱

J. Recoquilly 等研究了与鹌鹑行为和生产品特征相关的遗传图谱和 QTL。遗传图谱采用 2 145 个位于 28 个不同连锁群体的 1 479 个个体的 SNPs 构建而成,性别连锁图谱共跨越平均标记间距 2.1 cm 总长为 3 057 cm 的区域。除少数地区以外,日本鹌鹑采用跟日本鸡一样的标记顺序,连锁分析共揭示了 45 个与行为特性(23)或生产特性(22)相关的 QTLs,值得注意得是与社会动机特性相关的 QTL(15)最多。研究明确指出了控制情绪反应和鸟类体重(分别位于 *QTLCJA5* 和 *CJA8*)或社会行

为和开始产蛋(位于 QTLCJA19)的可能的多效性区域^[21]。S. Moon 等研究确定了猪在驯化过程中因人工选择导致激烈的表型变化相关的基因。对 30 头大白猪和大约克猪以及 10 头亚洲野猪进行全基因组重测序共获得了 19 990 个基因的 430 万个 SNPs。通过检测选择性清除构建了广泛的定向选择遗传图谱。研究表明在选择条件下的候选基因对猪的繁殖和生产有着重要的作用,定向选择的候选基因富集在影响大脑功能与饮食行为的谷氨酸代谢第 3 组受体上。其中基因 *ABLIM1*、*CXADR*、*INSR*、*RIMS1*、和 *SYNE1* 与生长调节有关;*BAI3*、*PKP4*、*PPFIA4* 和 *PCDHAC2* 与细胞粘附有关;*LIMS3*、*BAI3*、*CNTFR*、*PKP4* 和 *PCDHAC2* 与信号转导有关;*DNAJB5*、*ISOC1*、*METTL13*、*PPRC1* 和 *RBBP4* 与代谢有关。其中 *ZNF638* 是最值得关注的候选基因,它与编码早期脂肪形成调节相关的蛋白有关,并作为 *CEBPs* 辅助转录因子,控制 *PPARG* 的表达^[22]。

2 简化基因组测序

基于酶切的简化基因组测序是对与限制性核酸内切酶识别位点相关的 DNA 进行高通量测序,可大幅度降低基因组的复杂度,操作简便,同时不受参考基因组的限制,可快速鉴定出高密度的 SNP 位点,从而也能实现遗传进化分析及重要性状候选基因的预测。简化基因组技术也可广泛用于变异检测、遗传图谱构建、性状定位和群体进化研究。

2.1 群体进化

J. C. Jones 等利用 RAD-seq 测序技术研究了剑尾鱼属(花鱗科)包括 26 种来自中美洲的热带小型淡水鱼的系统发生关系,共检测了 143 个个体。测序深度 15X,共找到约 66 000 个 SNPs,构建了剑尾鱼属进化树,并推断了其祖先的剑尾形态^[23]。J. Gatchen 等利用 RAD-seq 研究了俄勒冈地区海水和淡水生境中三刺鱼的群体结构和定殖史。试验采用俄勒冈地区 9 个不同地点的三刺鱼 578 条,根据 SNP 位点,构建了鱼群的亲缘关系,分析了每个地点鱼群的遗传结构,Structure 结果推测 9 个地点的鱼群由 5 种鱼构成,分析结果表明,俄勒冈地区的三刺鱼是有记录的分化最早的三刺鱼,生活在该地区中部的三刺鱼是于近年人为进入定殖的^[24]。M. W. Jacobsen 等选取欧洲和美洲鳗鱼作为研究对象,研究了大西洋鳗基因组的印迹,通过 RAD 测序技

术,研究者们利用检测到的 328 300 SNPs,对欧洲鳗和美洲鳗两个姊妹物种的基因组印迹进行了阐述,在不考虑强烈分化的 3 757 个 ($F_{st} > 0.8$) SNPs 条件下,滑动窗口的 F_{st} 分析显示并没有高分化的基因组区域,总体 F_{st} 为 0.041,在分离出的大于 1 000 bp 的 SNPs 中几乎没有发现连锁不平衡,这反映了基因组搭便车的现象,物种之间的多个区域存在定向选择,GO 分析表明,包含正向选择的候选 SNP 的基因显著性富集于鳗的发展过程和磷酸化过程,这与幼虫阶段持续时间的差异和物种形成基础上的迁移距离的假设几乎一致,假定选择条件下的大多数 SNPs 于编码区外被发现,这为非编码区域的功能可能比之前假设的更重要这一新的观点提供了支持,总的来说,结果论证了在人口统计学参数和研究物种生活史特性结构下解释选择基因组印迹的必要性^[25]。L. Y. Rutledge 等采用 RAD-seq 技术研究了混血北美犬的起源,采用成年狼和幼狼已确定的 127 235 个 SNPs,结合基因组模拟物验证了东部北美洲各类犬的混合起源假说,主成分分析显示,并无证据证明东部地区的狼或者其他犬起源于灰狼和西部郊狼的杂交,研究结果支持东部狼作为一个独特的基因组类群存在于北美这一观点,而且也解决了大湖狼和东部郊狼的杂交起源^[26]。K. J. Emerson 等采用 RAD-seq 方法探索不同地理分布的瓶草蚊群体进化关系,对北美 21 个不同地理分布的瓶草蚊群体进行分析,采用部分碱基识别位点的内切酶 *Sbf* 1 消化 gDNA。在识别位点上共获得了 3 741 个 SNPs。使用最大似然法将 21 个群体区分为 4 个大分支,而且 21 个群体间的进化关系得到阐明^[27]。

2.2 变异检测

Z. Gompert 等采用简化多态序列复杂性 (CRoPS) 技术对北美红珠灰蝶 *Lycaeides* 的 12 个不同地域群体进行 SNP 标记检测,共获得 341 045 条序列,其中 36% 的序列变异在各群体中有分区现象^[28]。P. M. Richards 等采用 RAD-seq 技术研究了蜗牛属贝壳颜色和带型多态性,蜗牛属的贝壳颜色和带型基因型是由遗传决定的,从根本上来讲是由 ≥ 5 个基因座的超级基因决定的,研究发现,记录的 323 个个体颜色和条带在超级基因的 C 基因座和 B 基因座之间没有发现重组,其次对两个亲本、22 个后代使用 RAD-seq 技术,发现 44 个无特异性标记假设连接到颜色 (C) 基因座和带型 (B) 基因座。

RAD-SEQ 标记的最有可能的 11 个基因型在 22 个相同的后代中被独立验证,已标记的 C-B 超级连锁群中最近的 RAD-SEQ 位于 0.6 cM 范围内,与联合位点在一起构成了一个 38.5 cm 的连锁图^[29]。Z. Zhai 等首次利用 RAD-seq 技术研究了鸡的 SNP 和基因型分型,试验对 13 个中国本土鸡种和 3 个引进鸡种做了 RAD-seq,平均在每个个体中发现了 75 587 个 SNPs,在所有个体中进行严格筛选验证,得到了 28 895 个候选 SNPs,其中新发现了 15 404 个 SNPs^[30]。

2.3 性状定位

P. Andolfatto 等在拟果蝇上用低覆盖基因分型技术方法验证了在 8.5 kb 范围内存在性状决定基因,他们用表型为显性荧光眼基因型的拟果蝇家系同另外一种拟果蝇杂交,获得了 F2 个体,然后基于有参考基因组条件,采用 MSG 技术对其进行连锁分析,共获得 15 070 个用于标记的 SNPs,同时 QTL 分析发现 LOD 值与某一基因的距离十分相近^[31]。R. D. Houston 等从 20 个鱼群中选择 IPN 死亡率最高的 10 个群体应用微卫星标记进行 QTL 定位,对其中的 4 个亲本及其 7 个具有抗病性的纯合子和 7 个易染病的纯合子进行 RAD 测序,共获得了约 70 000 个酶切位点,检出 6 712 个 SNPs 标记位点,对两个家系全基因组 RAD 标记位点分离模式的分析表明,在 29 条染色体上均发现了 SNP 位点,且雄鱼的基因组缺乏重组,在所有的 SNPs 位点中,其中 50 个与 QTL 定位相关联,几个 SNPs 位点与 IPN 抗性显著性相关,且结果通过群体连锁不平衡 SNP 检测对其抗性进行了鉴定^[32]。

2.4 遗传图谱

B. K. Peterson 等利用双酶切系统的简化基因组测序技术在鹿鼠的两个姐妹物种 (*Maniculatus* 和 *Polionotus*) 的杂交群体中共获得了 1 000 多个 SNPs,且具有固定差异,并利用 1 000 多个 SNPs 构建了遗传图谱^[33]。C. Shao 等采用全基因组测序技术构建了比目鱼(牙鲆)的高分辨率遗传图谱,研究共鉴定了均匀分布在牙鲆基因组上的 13 362 个 SNPs,其中 12 712 高可信度的 SNPs 进行高通量基因分型,并分配到 24 个共识连锁群(LGs),遗传连锁图谱的总长度为 3 497.29 cm,其中基因座之间的平均距离为 0.47 cm,代表了当前日本比目鱼最密集的遗传图谱,位于遗传图谱上的大多数 LGs 和均匀分布的 3 类 SNPs(母系杂合、父系杂合和双杂合)

是饱和的,然而诸如 LGs(LG8、LG9、LG11、LG18 和 LG19)并没有,揭示了复杂的基因组包含地区丰富的重复序列和转座因子,有趣的是 LG8 包含的遗传距离最小,仅包含父系和双亲杂合的标记,并符合最小的物理距离。事实上,在母系图上 LG8 只有在双亲杂合标记,而在父系图上只有父系杂合和双杂合标记,表明这个 LG 上具有同源重组的特异性抑制^[34]。S. Gonen 等以两个大西洋鲑鱼群体为参考,利用 RAD-seq 技术研究了鲑鱼高密度 SNP 图谱,大约 60 000 个 SNPs 被分配到 29 个连锁群体中,然后分别构建了遗传连锁图谱,该连锁图谱与大西洋鲑鱼参考基因组草图重叠群进行了整合,其中 112 个基因组重叠群映射到两个或更多个连锁群体中,并强调了鲑鱼基因组中假定的部分同源配对的地区,与刺鱼参考基因组相比较进行基因组学分析得出假定的基因与大约一半的有序 SNPs 密切相关,并证明大西洋鲑鱼和刺鱼基因组之间存在直系同源性障碍,该研究为鲑鱼基因组学研究提供了遗传资源^[35]。S. W. Baxter 等首次对无参考基因组的小菜蛾进行简化基因组测序,通过对 24 个小菜蛾的个体进行 RAD 测序,将杀菌耐药性定位到 W/Z 性染色体上,并利用 RAD 等位基因构建了 1 个遗传连锁图谱,并获得了 31 个连锁群^[36]。P. M. Richards 等对雷默瑞丽蜗牛两个亲本和 22 个子代个体进行 RAD-seq 测序,在控制色带和色彩的基因座位上发现了 44 个标记位点,通过进一步的研究认证,最终构建了一张基于标记位点的连锁图谱,重新建立了雷默瑞丽蜗牛遗传模型^[37]。

3 重测序技术存在的问题

第二代高通量测序技术的不断改进,给重测序技术带来更新突破的同时也使得其局限性和挑战性日益凸显。以高通量定义的全基因组重测序技术和简化基因组测序技术往往会产生庞大的数据量,这会带来诸如数据储存和分析的困扰^[38]。首先,如何充分利用大数据是主要问题^[39];其次,如何经济有效地存储数据又是一个值得深思的难题^[40];最后,高通量测序所需的起始样本量可能更大,对于有限的样本的测序分析受到了限制^[41]。

如今以单分子读取技术为基础的第三代测序技术正在成为人们所关注的焦点^[42],由于三代测序技术是通过提高测序机器灵敏度以及增加荧光信号强度等手段来减小错误率和增大检出率,因此省去了

PCR 扩增这一步,数据读取速度更快更准,可直接测甲基化的 DNA 序列^[43-46],因此它的到来会更有利于这两种技术的应用。

从重测序技术的研究中可总结出几点问题:1)基因组测序产生的数据需精简;2)数据分析所用到的系统和软件普适性需待提高;3)操作产生的错误率有待降低,检出率有待提高;4)测序样本受限问题有待解决;5)测序经费有望更低。

4 展望

重测序技术涉及范围之广,其成绩毋庸置疑,对于大数据时代也是功不可没,虽然全基因组重测序技术和简化基因组测序技术并不是很完善,但作为新兴的高通量测序技术,其在基因组学研究方面已有举足轻重的地位。相信随着测序成本的逐步降低以及数据处理和分析方法的不断完善,它们必将在未来基因组学的研究中起重要作用。未来的发展探究中,可能会更注重数据挖掘,个性化分析,以及动物基因改良、疾病预防、治疗等。

参考文献(References):

- [1] BENTLEY D R. Whole-genome re-sequencing [J]. *Curr Opin Genet Dev*, 2006, 16(6): 545-552.
- [2] DAVEY J W, HOHENLOHE P A, ETTER P D, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing [J]. *Nat Rev Genet*, 2011, 12(7): 499-510.
- [3] GAO Q, YUE G, LI W, et al. Recent progress using high-throughput sequencing technologies in plant molecular breeding [J]. *J Integr Plant Biol*, 2012, 54: 215-227.
- [4] GOFF S A, RICKE D, LAN T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*) [J]. *Science*, 2002, 296(5565): 92-100.
- [5] FUJISAWA M, BABA T, NAGAMURA Y, et al. The map-based sequence of the rice genome [J]. *Nature*, 2005, 436(436): 793-800.
- [6] QU Y, ZHAO H, HAN N, et al. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau [J]. *Nat Commun*, 2013, 4(1): 14.
- [7] LI M, TIAN S, JIN L, et al. Genome analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars [J]. *Nat Genet*, 2013, 45(12): 1431-1438.
- [8] XIA Q, GUO Y, ZHANG Z, et al. Complete resequencing of 40 genomes reveals domestication events and genes in Silkworm (*Bombyx*) [J]. *Science*, 2009, 326(5951): 433-436.
- [9] ZHOU X, WANG B, PAN Q, et al. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history [J]. *Nat Genet*, 2014, 46(12): 1303-1310.
- [10] ZHANG R, YIN Y, ZHANG Y, et al. Molecular characterization of transgene integration by next-generation sequencing in transgenic cattle [J]. *PLoS One*, 2012, 7(11): e50348.
- [11] SEO D W, OH J D, JIN S, et al. Single nucleotide polymorphism analysis of Korean native chickens using next generation sequencing data [J]. *Mol Biol Rep*, 2015, 42(2): 471-477.
- [12] KEIGHTLEY P D, TRIVEDI U, THOMSON M, et al. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines [J]. *Genome Res*, 2009, 19(7): 1195-1201.
- [13] RUBIN C J, ZODY M C, ERISKSSON J, et al. Whole-genome resequencing reveals loci under selection during chicken domestication [J]. *Nature*, 2010, 464(7288): 587-594.
- [14] LEE K T, CHUNG W H, LEE S Y, et al. Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity [J]. *BMC Genomics*, 2013, 14(2): 373-379.
- [15] ZHAN B, FADISTA J, THOMSEN B, et al. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping [J]. *BMC Genomics*, 2011, 12(71): 16107-16112.
- [16] YI G, QU L, LIU J, et al. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing [J]. *BMC Genomics*, 2014, 15(1): 1-16.
- [17] AXELSSON E, RATNAKUMAR A, ARENDT M L, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet [J]. *Nature*, 2013, 495(7441): 360-364.
- [18] BOWEN M E, HENKE K, SIEGFRIED K, et al. Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing [J]. *Genetics*, 2011, 190(3): 1017-1024.
- [19] MERKER J D, ROSKIN K M, NG D, et al. Comprehensive whole-genome sequencing of an early-stage primary myelofibrosis patient defines low mutational burden and non-recurrent candidate genes [J]. *Haematologica*, 2013, 98(11): 1689-1696.
- [20] DAETWYLER H D, CAPITAN A, PAUSCH H, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle [J]. *Nat Genet*, 2014, 46(8): 858-865.
- [21] RECOQUILLAY J, PITEL F, ARNOULD C, et al. A medium density genetic map and QTL for behavioral and production traits in Japanese quail [J]. *BMC Ge-*

- nomics, 2015, 16(1):1-12.
- [22] MOON S, KIM T H, LEE K T, et al. A genome-wide scan for signatures of directional selection in domesticated pigs[J]. *BMC Genomics*, 2015, 16(1):1-12.
- [23] JONES J C, FAN S, FRANCHINI P, et al. The evolutionary history of Xiphophorus fish and their sexually selected sword; a genome-wide approach using restriction site-associated DNA sequencing [J]. *Mol Ecol*, 2013, 22(11):2986-3001.
- [24] GATCHEN J, BASSHAM S, WILSON T, et al. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing [J]. *Mol Ecol*, 2013, 22(11):2864-2883.
- [25] JACOBSEN M W, PUJOLAR J M, BERNATCHEZ L, et al. Genomic footprints of speciation in Atlantic eels (*Anguilla anguilla* and *A. rostrata*) [J]. *Mol Ecol*, 2014, 23(19):4785-4798.
- [26] RUTLEDGE L Y, DEVILLARD S, BOONE J Q, et al. RAD sequencing and genomic simulations resolve hybrid origins within North American Canis[J]. *Biol Lett*, 2015, 11(7). pii: 20150303. doi: 10. 1098/rsbl. 2015. 0303.
- [27] EMERSON K J, MERZ C R, CATCHEN J M, et al. Resolving postglacial phylogeography using high-throughput sequencing[J]. *Proc Natl Acad Sci U S A*, 2010, 107(37):16196-16200.
- [28] GOMPERT Z, FORISTER M L, FORDYCE J A, et al. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genom of Lycaeides butterflies [J]. *Mol Ecol*, 2010, 19(12):2455-2473.
- [29] RICHARDS P M, LIU M M, LOWE N, et al. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene[J]. *Mol Ecol*, 2013, 22(11):3077-3089.
- [30] ZHAI Z, ZHAO W, HE C, et al. SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens[J]. *Anim Genet*, 2015, 46(2):216-219.
- [31] ANDOLFATTO P, DAVISON D, EREZYILMAZ D, et al. Multiplexed shotgun genotyping for rapid and efficient genetic mapping[J]. *Genome Res*, 2011, 21(4):610-617.
- [32] HOUSTON R D, DAVEY J W, BISHOP S C, et al. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon[J]. *BMC Genomics*, 2012, 13(12):1-15.
- [33] PETERSON B K, WEBER J N, KAY E H, et al. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species[J]. *PLoS One*, 2012, 7(5):e37135.
- [34] SHAO C, NIU Y, RASTAS P, et al. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis[J]. *DNA Res*, 2015, 22(2):161-170.
- [35] GONEN S, LOWE N R, CEZARD T, et al. Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing [J]. *BMC Genomics*, 2014, 15(3):166.
- [36] BAXTER S W, DAVEY J W, JOHNSTON J S, et al. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism[J]. *PLoS One*, 2011, 6(4):e19315.
- [37] RICHARDS P M, LIU M M, LOWE N, et al. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene[J]. *Mol Ecol*, 2013, 22(11):3077-3089.
- [38] 黎 裕, 李英慧, 杨庆文, 等. 基于基因组学的作物种质资源研究: 现状与展望[J]. *中国农业科学*, 2015, 48(17):3333-3353.
- LI Y, LI Y H, YANG Q W, et al. Research on crop germplasm resources based on genomics: present situation and prospect[J]. *Agricultural Sciences in China*, 2015, 48(17):3333-3353. (in Chinese)
- [39] VAN VLIET A H. Next generation sequencing of microbial transcriptomes: challenges and opportunities [J]. *FEMS Microbiol Lett*, 2010, 302(1):1-7.
- [40] BAKER M. Gene data to hit milestone[J]. *Nature*, 2012, 487(7407):282-283.
- [41] CHEN H, TAN X F. Excavation of genic resources based on next generation sequencing technologies[J]. *Plant Physiol J*, 2014, 50(8):1089-1095.
- [42] EID J, FEHR A, GRAY J, et al. Real-time DNA sequencing from single polymerase molecules[J]. *Science*, 2009, 323(5910):133-138.
- [43] LEVEBE M J, KORLACH J, TURNER S W, et al. Zero-mode waveguides for single-molecule analysis at high concentrations [J]. *Science*, 2003, 299(5607):682-686.
- [44] HARRIS T D, BUZBY P R, BABCOCK H, et al. Single-molecule DNA sequencing of a viral genome[J]. *Science*, 2008, 320(5872):106-109.
- [45] TOMBACZ D, SHARON D, OLAH P, et al. Strain Kaplan of pseudorabies virus genome sequenced by pacBio single-molecule real-time sequencing technology[J]. *Genome Announc*, 2014, 2(4):e006028.
- [46] STODDART D, HERON A J, MIKHAILOVA E, et al. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore[J]. *Proc Natl Acad Sci U S A*, 2009, 106(19):7702-7707.